

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 03:18:06 ; Search time 8.96557 Seconds  
(without alignments)  
1581.285 Million cell updates/sec

Title: US-09-807-933b-1

Sequence: 1 MKFTTASALLALALGTEM.....TYKEVTCPEKITAKTCSRK 338

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	37.8	376	1 GUNK_FUSOX	P45699 fusarium ox
2	679	37.0	213	1 GUNK_HOMIN	P43316 humicola in
3	552.5	30.1	511	1 GUNK_PSEFL	P18126 pseudomonas
4	515.5	28.1	393	1 GUNK_USDMA	P54424 usdago ma
5	241	13.1	471	1 GUNK_TRIRE	P07987 trichoderma
6	180	9.8	418	1 GUNK_TRIRE	P07982 trichoderma
7	179	9.7	462	1 GUNK_FUSOX	P46236 fusarium ox
8	173	9.4	438	1 GUNK_AGABI	P49075 agaricus bi
9	153.5	8.4	385	1 GUNK_FUSOX	P46329 fusarium ox
10	150	8.2	2704	1 G168_PAPPR	P17053 paramedium
11	149.5	8.1	475	1 SIMI_YEAST	P40472 saccharomyc
12	148.5	8.1	540	1 GUNK_ASPAC	O59843 aspergillus
13	148	8.1	210	1 PSBE_PORPU	P50272 porphyra pu
14	147.5	8.0	442	1 GUNK_DICDI	P54639 dictyostell
15	147	8.0	242	1 GUNK_TRIRE	P43317 trichoderma
16	144	7.8	2715	1 G156_PAPPR	P13390 bacterioph
17	143	7.8	1396	1 VTRF_BPT5	P10476 pseudomonas
18	141	7.7	962	1 GUNK_PSEFL	O83172 treponema p
19	138	7.5	485	1 Y136_TREPA	O92400 agaricus bi
20	138	7.5	506	1 GUNK_AGABI	P47033 saccharomyc
21	135.5	7.4	881	1 YVH8_YEAST	O12624 humicola in
22	135	7.4	388	1 GUNK_HOMIN	P00725 trichoderma
23	133	7.2	513	1 GUNK_TRIRE	P19355 trichoderma
24	132	7.2	513	1 GUNK_TRIRE	P54643 dictyostell
25	130	7.1	677	1 SP87_DICDI	P92127 giardia lam
26	129.5	7.1	687	1 VS41_GIALA	P23490 homo sapien
27	129	7.0	316	1 LORI_HUMAN	P07856 bombyx mori
28	128.5	6.9	389	1 SERI_BOMMO	O01838 listeria se
29	127.5	6.9	523	1 P60_LISSE	P15269 dictyostell
30	127.5	6.9	537	1 SP76_DICDI	P46238 fusarium ox
31	126	6.9	514	1 GUNK_FUSOX	P78504 homo sapien
32	125	6.8	1218	1 JAGI_HUMAN	P53616 saccharomyc
33	124.5	6.8	420	1 SUN4_YEAST	

34	124.5	6.8	459	1 GUNK_TRIRE	P07981 trichoderma
35	124	6.8	1218	1 JAGI_MOUSE	O9qxx0 mus musculu
36	123.5	6.7	937	1 HYR1_CANAL	P46591 candida alb
37	123.5	6.7	1140	1 YN66_YEAST	O04893 saccharomyc
38	122.5	6.7	450	1 UTH1_YEAST	P36135 saccharomyc
39	122.5	6.7	1210	1 ICEN_PSEFL	P09815 pseudomonas
40	121.5	6.6	604	1 YN23_YEAST	P53632 saccharomyc
41	121.5	6.6	503	1 AMYG_RHIOH	P07683 rhizopus or
42	121	6.6	537	1 GUNK_PENNA	O06886 penicillium
43	121	6.6	1150	1 APNU_PIG	P12021 sus scrofa
44	120.5	6.6	1196	1 ICEV_PSEEX	O33479 pseudomonas
45	120	6.5	320	1 CEL1_AGABI	O00023 agaricus bi

## ALIGNMENTS

RESULT 1	ID	GUNK_FUSOX	STANDARD;	PRT;	376 AA.
AC	P45699;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	Putative endoglucanase type K precursor (EC 3.2.1.4) (Endo-1,4-Beta-glucanase) (Cellulase).				
OS	Fusarium oxysporum.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OX	Hypocreales; Mitosporic Hypocreales; Fusarium.				
OX	NCBI_TaxID=5507;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=95047531; PubMed=7959045;				
RA	Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,				
RA	Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;				
RT	"The use of conserved cellulase family-specific sequences to clone				
RT	cellulase homologue cDNAs from Fusarium oxysporum."				
RL	Gene 150:163-167(1994).				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic				
CC	linkages in cellulose.				
CC	-1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).				
CC	-1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL				
CC	HYDROLASES).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
CC	EMBL; I29381; AAA6589.1; -				
DR	HSSP; P43316; 2BNG.				
DR	InterPro; IPR000254; CBD_fungal.				
DR	InterPro; IPR000334; GH_45.				
DR	Pfam; PF00734; CBM_1; 1_				
DR	Pfam; PF02015; GLYCO_hydro_45; 1.				
DR	SMART; SM00236; ICBID; 1.				
DR	PROSITE; PS00562; CBD_FUNGAL; 1.				
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.				
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.				
FT	SIGNAL	1	16	POTENTIAL.	
FT	CHAIN	19	376	POTATIVE ENDOGLUCANASE TYPE K.	
FT	DOMAIN	19	308	CATALYTIC.	
FT	DOMAIN	309	338	LINKER.	
FT	DOMAIN	339	376	CELLULOSE-BINDING.	
FT	ACT SITE	29	29	NECLOPHILE (BY SIMILARITY).	
FT	ACT SITE	140	140	PROTON DONOR (BY SIMILARITY).	
SQ	SEQUENCE	376 AA.	39235 MM;	B430A5F9628B9882 CRC64;	

Query Match 37.8%; Score 694; DB 1; Length 376;  
Best Local Similarity 58.6%; Pred. No. 3.2e-39;

Matches 126; Conservative 33; Mismatches 48; Indels 8; Gaps 5;

QY 124 AVSGAGSNGVTRRYDCCKASCSMPGKANVSPVSKCNKDGVTALSDSNAGCNGNS 183  
 DB 14 AVS-AASGSHSRMYDCCKPCSCSMGKAHVAPALTCND-NPISNTNVAVNGCEGGS 71  
 QY 184 -YACNDNQPANVNDNLAYGFAAALISGGESRWCCSCFELTFTSTSVAGKXVVOVTNTG 242  
 DB 72 AYACTNTPSPAVNDELAYGFAATKISGSEASWCCACTALTTTGPKKMTVOGTNTG 131  
 QY 243 GDGSSSTGAFHDLQMPGCGVIGFNGCSSQWGAHPNDGWSRYGISSASDCSSLPALQAG 302  
 DB 132 GDLDGN---HFGLMPPGGVGIGFDGCTSEFGKLG--GAQYGISRSRSCDSYPELLKDG 186  
 QY 303 CKMRFNFPKANDNPSMTYKCVTKPKETATKTCGR 337  
 DB 187 CHMRFDFWENADNDPFTFEQVOCPEKALLDISGCKR 221

## RESULT 2

ID GUNB HUMIN STANDARD; PRT; 213 AA.

AC P4316;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)  
 DE (Cellulase V) (Bg V).  
 OS Humicola insolens.  
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.  
 NCBI\_TaxID=34413;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rasmussen G., Mikkelsen J.-M., Schulein M., Patkar S.A., Hagen F.,  
 RA Hjort C.M., Hestrup S.;  
 RT "A cellulase preparation comprising an endoglucanase enzyme.";  
 RL Patent number WO9117243, 14-NOV-1991.  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RX MEDLINE=93390621; PubMed=8377830;  
 RA Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z.,  
 RA Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schulein M.;  
 RT "Structure and function of endoglucanase V.";  
 RL Nature 365:362-364 (1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RX MEDLINE=96101453; PubMed=8519779;  
 RA Davies G.J., Tolley S.P., Henriksen B., Hjort C., Schulein M.;  
 RT "Structures of oligosaccharide-bound forms of the endoglucanase V  
 from Humicola insolens at 1.9-A resolution.";  
 RL Biochemistry 34:16210-16220 (1995).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).  
 RA Davies G.J., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z.,  
 RA Wilson K.S., Rasmussen G., Schulein M.;  
 RT "Structure determination and refinement of the Humicola insolens  
 endoglucanase V at 1.5-A resolution.";  
 RL Acta Crystallogr. D 52:7-17 (1996).  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 linkages in cellulose.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL  
 HYDROLASES).  
 CC PDB; 2ENG; 08-DEC-96.  
 CC PDB; 3ENG; 16-JUN-97.  
 CC PDB; 4ENG; 16-JUN-97.  
 DR InterPro; IPR000334; GH\_45.  
 DR Pfam; PF02015; Glyco\_hydro\_45; 1.  
 DR PROSITE; PS01140; GLYCOSYL\_HYDROL\_45; 1.  
 KM Cellulose degradation; Hydrolysis; Glycosidase; 3D-structure.  
 FT ACT\_SITE 10  
 FT ACT\_SITE 10 NUCLEOPHILE.  
 FT PROTON\_DONOR.  
 SQ SEQUENCE 213 AA; 22864 MW; 24334301BA3BC04 CRC64;

Query Match 37.0%; Score 679; DB 1; Length 213;  
 Best Local Similarity 59.1%; Pred. No. 1.8e-38;  
 Matches 123; Conservative 31; Mismatches 44; Indels 10; Gaps 5;

QY 132 NGVTRRYDCCKACSMGKANVSPVSKCNKDDVTALSDSNAGCNGNS 190  
 DB 2 DGRSTRYDCCKPCSCSMGKAHVAPALTCND-NPISNTNVAVNGCEGGS 60  
 QY 191 PAVANDNLAYGFAAALISGGESRWCCSCFELTFTSTSVAGKXVVOVTNTGDLGSSSTG 250  
 DB 61 PAVAVDDDLALGFAANISGNEAGWCCACTALTTTGPKKMTVOGTNTG 118  
 QY 251 AHFDLQMPGCGVIGFNGCSSQWGA-PNDGWSRYGISSASDCSSLPALQAGCKMRFN 309  
 DB 119 -HFGLNIPGGVGIGFDGCTPQFGGLP---GQRYGISRSRSCDSYPELLKDG 173  
 QY 310 FKNADNPSMTYKCVTKPKETATKTCGR 337  
 DB 174 FKNADNPSFSPRQVCPALVARTGCR 201

## RESULT 3

ID GUNB\_PSEFL STANDARD; PRT; 511 AA.

AC P18126;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase) (ECB).  
 GN ECB.  
 OS Pseudomonas fluorescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 NCBI\_TaxID=294;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.  
 RC STRAIN=SP. Cellulosa;  
 RX MEDLINE=90355836; PubMed=2117693;  
 RA Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;  
 RT "The N-terminal region of an endoglucanase from Pseudomonas  
 fluorescens subspecies cellulosa constitutes a cellulose-binding  
 domain that is distinct from the catalytic centre.";  
 RL Mol. Microbiol. 4:759-767 (1990).  
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-  
 GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
 GLUCANS. ECB IS MOST ACTIVE AGAINST AMORPHOUS AND CRYSTALLINE CELLULOSE.  
 CC SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRYSTALLINE CELLULOSE.  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 linkages in cellulose.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN  
 (CBD).  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL  
 HYDROLASES).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL collaboration  
 between the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X52615; CA36844.1; --  
 CC PIR; S10527; S10527.  
 DR HSSP; P43316; 2ENG.  
 DR InterPro; IPR001919; Bac cellose-bind.  
 DR InterPro; IPR002883; CBD\_5.  
 DR InterPro; IPR000334; GH\_45.  
 DR Pfam; PF00553; CBM\_2; 1.  
 DR Pfam; PF02013; CBM\_10; 1.  
 DR Pfam; PF02015; Glyco\_hydro\_45; 1.

DR PROSITE; PS00561; CBD BACTERIAL; 1.  
 DR PROSITE; PS01140; GLYCOSYL HYDROL P45; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Periplasmic.  
 FT SIGNAL 1 29  
 FT CHAIN 30 511  
 FT DOMAIN 30 131  
 FT DOMAIN 132 173  
 FT DOMAIN 223 259  
 FT DISULFID 32 127  
 FT ACT SITE 276 276  
 FT ACT SITE 393 393  
 SQ SEQUENCE 511 AA; 52078 MW; 3C3119998291D8E CRC64;

Query Match 30.1%; Score 552.5; DB 1; Length 511;  
 Best Local Similarity 35.2%; Pred. No. 9.4e-30;  
 Matches 146; Conservative 48; Mismatches 128; Indels 93; Gaps 16;

QY 4 ITTASALLALALGTEMASAECSKLYQ-CGKKNMNGPTCCS-----GSTCKVAN 54  
 DB IPIGSSVERGVQ-GNNGSRAQVPAVTGACGGSSAPSSVSSSSSVSTPRSSS 156  
 QY 55 DRYSQLPGSSGNKSESNAHK-----TTAAHKTTT-----AAHKTTAP 98  
 DB 157 SSVSSSVPTSSSSSVLTGAQCNWYGTLPICNNTSNGWGEDGRSCVARTTSAOP 216  
 QY 99 AKKTTTAAKASTNSSS-----SSSGKSAVSGASG-----NGVTRRYMD 140  
 DB 217 APYGIYSTSTSTLSSSSSSSSSVASSSSLSATSSASVSVPPIDGCKNGATRYMD 276  
 QY 141 CCKASCSWPKKANVS---SPVKSCKNDGVTALSDNAQSCNGNSYMCNDNPMAVNDN 197  
 DB 277 CCKPHCGM--SANNVPLSVPLQSCANN--TRLSADVSGSCDGGGVMCDKIPAVASPT 333  
 QY 198 LAYGFAAAISGGESBWCSCPELTFT-----STVAGKKNVQVNTTGGDLG 246  
 DB 334 LAYGVAATG-SGG-----VCCRCYLOFTGSSSYNAPGDPGSAALAGKMTVOATNIGYDV- 387  
 QY 247 SSTGAPHDLOMPGGGVGIFNGCSSOMGAPNDGMSRGYGGISAA-----289  
 DB 388 --SGGQFDILVPGGAGAFACSAQMSVNAELQAGGFLACKQDLGYNASLSQYKSC 445  
 QY 290 --SDCSSL-----PSALQAGCKRFFWFKNADNPMTYKEVTCPKEITTAQTGCR 337  
 DB 446 VLMRCDSVFESRGLTQDQGCPTFAEFEAADNPULKYKEVPPAELTTRSGNMR 500

RESULT 4  
 GUNI\_USTWA STANDARD; PRT; 393 AA.

AC P54424;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)  
 DE (Cellulase 1) (EG 1).  
 GN EGL1.  
 OS Ustilago maydis (Smut fungus).  
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;  
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
 OX NCBI\_TaxID=5270;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PD11;  
 RX MEDLINE=96145728; PubMed=8590631;  
 RA Schaeffer F., Wanner G., Kahmann R.;  
 RT "Filament-specific expression of a cellulase gene in the dimorphic  
 fungus Ustilago maydis";  
 RL Biol. Chem. Hoppe-Seyler 376:617-625 (1995).  
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 linkages in cellulose.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: HYPMAL TIP.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN FILAMENTOUS DIKARYON.

CC -1- PTM: MAY ALSO BE O-GLYCOSYLATED.  
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL  
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; S81598; AAB36147.1; -

DR HSSP; P43316; ZENG.

DR InterPro: IPR000334; GH 45.

DR Pfam: PF02015; Glyco\_hydro\_45; 1.

DR PROSITE; PS01140; GLYCOSYL HYDROL P45; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Glycoprotein.

FT SIGNAL 1 26  
 FT CHAIN 27 393  
 FT ACT SITE 34 34  
 FT ACT SITE 152 152  
 FT ACT SITE 270 270  
 FT DOMAIN 343 343  
 FT CARBOHYD 343 343  
 SQ SEQUENCE 393 AA; 39594 MW; 65C753C610CD6AD3 CRC64;

Query Match 28.1%; Score 515.5; DB 1; Length 393;  
 Best Local Similarity 47.0%; Pred. No. 2e-27; Indels 21; Gaps 9;  
 Matches 103; Conservative 31; Mismatches 64;

QY 133 GVTTRYWDCCKASCSWPKKANVS-SPVKSCKNDGVTAL---SDSNAGCNGNSYMCNDN 189  
 DB 27 GMATRYWDCCLASASMEGAPYAPVADCKADGVTILIDSKDPSGSGCNGNKKPFCSGM 86  
 QY 190 QPW--AVNDNLAYGFAAAISGGESBWCSCPELTFT---TSTVAGKKNVQVNTTGG 243  
 DB 87 QPFDDEDTPLAFGF--GAFITGGQSDTDCAFYAEFEHDAQKAMKRNKLI.FQVTVNNG 144  
 QY 244 DLGSSGAPHDLOMPGGGVGIF-NGCSSOMGAPNDGMSRGYGGISASPSLQNG 302  
 DB 145 DVQSQ---NFDQIPGCGIGAPFKCPCPAQGVASLWQGVASVATSCSLPRLQRC 201  
 QY 303 CKMRFN-WFKNADNPMSM--TYKEVTCPKEITTAQTGCRK 338  
 DB 202 CKMRSEW---GDNFVLKGSPPKRVKCPKSLIDRSGCRK 237

RESULT 5  
 GUX2\_TRIRE STANDARD; PRT; 471 AA.

AC P07987;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellulohydrolyase II)  
 DE (CBHII) (1,4-beta-cellulohydrolyase).  
 GN CBH2.  
 OS Trichoderma reesei (Hypocrea jecorina).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; Hypocreaceae; Hypocrea.  
 OX NCBI\_TaxID=51453;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VT-D-80133;  
 RX MEDLINE=87248061; PubMed=3596237;  
 RA Teeri T.T., Lehtovaara P., Kauppinen S., Salovuori I., Knowles J.;  
 RT "Homologous domains in Trichoderma reesei cellulolytic enzymes: gene  
 RT sequence and expression of cellobiohydrolyase II.";  
 RL Gene 51:43-52 (1987).  
 CC (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Qm9414 / Rut C-30;  
 RA Chen C.M., Gritzall M., Stafford D.W.;

RT "Nucleotide sequence and deduced primary structure of  
 RT cellobiohydrolase II from *Trichoderma reesei*,"  
 RL Biotechnology 5:274-278(1987).  
 RN [3]  
 RP SEQUENCE OF 25-44.  
 RA Faegerstaem L.G., Petersson L.G.  
 RT "The 1,4-beta-glucan cellobiohydrolases of *Trichoderma reesei* QM  
 RT 9414,"  
 RL FEBS Lett. 119:97-100(1980).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=90333255; PubMed=2377893.  
 RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.,  
 RT "Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*,"  
 RL Science 249:380-386(1990).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=97029636; PubMed=8875646.  
 RA Koivula A., Reihkainen T., Ruohonen L., Valkenjaervi A.,  
 RA Claessens M., Telenan O., Kleywegt G.J., Szardenings M., Rouvinen J.,  
 RA Jones T.A., Teeri T.T.  
 RT "The active site of *Trichoderma reesei* cellobiohydrolase II: the role  
 RT of tyrosine 169,"  
 RL Protein Eng. 9:691-699(1996).  
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE.  
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:  
 CC (1) ENDOCYLICANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;  
 CC (2) EXOCYLICANASES THAT CUT THE DISACCHARIDE CELLULOSE  
 CC FROM THE NONREDUCTING END OF THE CELLULOSE POLYMER CHAIN;  
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER  
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-Beta-D-glucosidic linkages  
 CC in cellulose and cellobiose, releasing cellobiose from the non-  
 CC reducing ends of the chains.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: T RESEI PRODUCES TWO DIFFERENT  
 CC EXOCYLICANASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE  
 CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOCYLICANASES.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL  
 CC HYDROLASES).  
 CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation --  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).  
 CC or send an email to [license@ib-sib.ch](mailto:license@ib-sib.ch)).  
 CC -----  
 DR EMBL, M16190, AAA34210.1; -.  
 DR EMBL, M5080, AAA72922.1; -.  
 DR PIR, A26472, A26472.  
 DR PIR, A26160, A26160.  
 DR PDB, 3CBH; 15-JAN-91.  
 DR PDB, 1CB2; 19-MAR-99.  
 DR InterPro, IPR000254; CBD fungal.  
 DR InterPro, IPR001524; GH\_6.  
 DR Pfam, PF00734; CBM\_1; 1.  
 DR Pfam, PF01341; Glyco\_hydro\_6; 1.  
 DR PRINTS, PR00733; GLHYDRASE6.  
 DR ProDom, PD001821; CBD fungal; 1.  
 DR ProDom, PD003733; GH\_6; 1.  
 DR SMART, SM00236; fCBM\_1.  
 DR PROSITE, PS00562; CBD\_FUNGAL; 1.  
 DR PROSITE, PS00655; GLYCOSYL\_HYDROL\_F6\_1; 1.  
 DR PROSITE, PS00656; GLYCOSYL\_HYDROL\_F6\_2; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;  
 KW 3D-structure.  
 FT SIGNAL 1 24  
 FT CHAIN 25 471 EXOGLUCANASE II.  
 FT DOMAIN 25 65 CELLULOSE-BINDING (BY SIMILARITY).

FT	DOMAIN	66	107	.471	LINKER.
FT	ACT_SITE	199	199		CATALYTIC.
FT	ACT_SITE	245	245		PROTON DONOR.
FT	ACT_SITE	425	425		NUCLEOPHILE.
FT	DISULFID	34	51		BY SIMILARITY.
FT	DISULFID	45	61		BY SIMILARITY.
FT	DISULFID	200	259		
FT	DISULFID	392	439		
FT	CARBOHYD	38	38		N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	111	111		O-LINKED (MAN. .)
FT	CARBOHYD	121	121		O-LINKED (MAN. .)
FT	CARBOHYD	130	130		O-LINKED (MAN. .)
FT	CARBOHYD	133	133		O-LINKED (MAN. .)
FT	CARBOHYD	134	134		O-LINKED (MAN. .)
FT	CARBOHYD	139	139		O-LINKED (MAN. .)
FT	CARBOHYD	313	313		N-LINKED (GLCNAC. .)
FT	CARBOHYD	334	334		N-LINKED (GLCNAC. .)
FT	MUTAGEN	199	199		D->A. 20% OF WILD-TYPE ACTIVITY.
FT	MUTAGEN	245	245		D->A. NO MEASURABLE ACTIVITY.
FT	CONFLICT	359	359		P -> R (IN REF. 2).
FT	CONFLICT	449	449		P -> A (IN REF. 2).
SO	SEQUENCE	471 AA;	49653 MW;		CA711BC35B1BD88 CRC64;
<hr/>					
Query Match					
Best Local Similarity 40.5%; Pred. No. 3.2e-09; Length 471;					
Matches 51; Conservative 21; Mismatches 46; Indels 8; Gaps 3					
Oy	5	TIASSLLALATGEMASAECSKLYGGCGGNMNGPTCCSGSTCKVSNDDYYSQCLPSG	64		
Db	8	TLATLTATLASVPLEBRQA--CSSWVGCGGGGNNMGPTCCAGSSTCYVSNDDYYSQCLPGA	65		
Oy	65	SSGNKSESNAHKKTTHAKHTTTPAKKTTTYAKASTPSNSSSSSGKYSA	124		
Db	66	ASSSSSTRPA---SITSRVSPETTRSSSATPPPGSTTRVPPVG--SGTATYSGNPFVG	119		
Oy	125	VSGGAS	130		
Db	120	VTFPMAN	125		
<hr/>					
RESULT 6					
GUN2_TRIRE	ID	GUN2_TRIRE	STANDARD;	PRT;	418 AA.
AC	P07982;				
DT	01-AUG-1988	(Rel. 08,	Created)		
DT	01-AUG-1988	(Rel. 08,	Last sequence update)		
DT	15-JUL-1999	(Rel. 38,	Last annotation update)		
DE	Endoglucanase EG-II precursor	(EC 3.2.1.4)	(Endo-1,4-beta-glucanase)		
DE	(Cellulase).				
GN	EGII OR EGI.II.				
OS	Trichoderma reesei (Hypocrea jecorina).				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OC	Hypocreales; Hypocreaceae; Hypocrea.				
OX	NCBI_TaxID=51453;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=VTT-D-80133;				
RX	MEDLINE=88255850; PubMed=3384334;				
RA	Saloheimo M., Lehtovaara P., Penttilae M., Teeri T.T., Staahlberg J.,				
RA	Johansson G., Pettersson G., Clayssens M., Tomme P., Knowles J.K.C.;				
RT	"EgIII, a new endoglucanase from Trichoderma reesei: the				
RT	characterization of both gene and enzyme.";				
RL	Gene 63:11-21(1988).				
RN	[2]				
RP	ACTIVE SITE GLU-350.				
RX	MEDLINE=93131031; PubMed=8093602;				
RA	Macarion R., van Beeumen J., Henriksat B., de la Mata I.,				
RA	Claessens M.;				
RT	"Identification of an essential glutamate residue in the active site				
RT	of endoglucanase III from Trichoderma reesei.";				
FEBS Lett.	316:137-140(1993).				
-!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE					

```

CC CC GENERAL REQUIREMENTS THREE TYPES OF HYDROLYTIC ENZYMES:
CC CC (1) ENDOGYUCANASES WHICH CUT INTERNAL BETA-1,4-GLYCOSIDIC BONDS;
CC CC (2) EXOCELLULOSEHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
CC CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYSE THE CELLULOSE AND OTHER
CC CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC CC linkages in cellulose.
CC CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC CC HYDROLASES) .
CC CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD) .
CC CC -1- CAUTION: WAS ORIGINALLY CALLED ENDOGYUCANASE EG-II .
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-eb.ch/announce/
CC CC or send an email to license@isb-eb.ch) .
CC CC -----
CC CC EMBL: M19373; AAA34213.1; .
CC CC DR PIR: S28372; S28372.
CC CC DR HSSP: P00725; 2CBH.
CC CC DR InterPro: IPR000254; CBD_fungal.
CC CC DR InterPro: IPR01547; GH_5.
CC CC DR Pfam: PF00150; cellulase; 1.
CC CC DR Pfam: PF00734; CBM 1; 1.
CC CC DR Prodom: PD001821; CBD_fungal; 1.
CC CC DR SMART: SMO0236; fCBD_1.
CC CC DR PROSITE: PS00562; CBD_FUNGAL; 1.
CC CC DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
CC CC DR KMW Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; signal.
CC CC FT SIGNAL 1 21
CC CC FT CHAIN 22 418 ENDOGLUCANASE EG-II.
CC CC FT DOMAIN 22 57 CELLULOSE-BINDING (BY SIMILARITY) .
CC CC FT DOMAIN 58 91 LINKER.
CC CC FT MOD RES 92 418 CATALYTIC.
CC CC FT CARBOHYD 124 124 PYRROLIDONE CARBOXYLIC ACID.
CC CC FT DISULFID 29 46 N-LINKED (GLCNAC. . .) (POTENTIAL) .
CC CC FT DISULFID 40 56 BY SIMILARITY.
CC CC FT ACT_SITE 239 239 PROTON DONOR (BY SIMILARITY) .
CC CC FT ACT_SITE 350, 350 NUCLEOPHILE.
CC CC SQ SEQUENCE 418 AA; 44227 MW; 26A492D55237A49B CRC64;

Query Match 9.8%; Score 180; DB 1; Length 418;
Best Local Similarity 23.5%; Pred. No. 3e-05; Indels 76; Gaps 10.
Matches 68; Conservative 36; Mismatches 109;

QY 8 SSALLALALGTEMASAECSKLYGCGGKMNWPTCCSGSTCKYSNDYSCQLPSCGSG 67
DB 4 SVALPULLAASILYGAVAQVQYTWGCGGIGMSGPNCAFGSACSTLNYIYACIDGATT 62
QY 68 NKSESESAHKTTTAAHKTTTAAHKTTTAAKKTTPAAKSTPNSSSSSSG- KYSAVS 126
DB 63 ----- IIT----- STRPSGPTTTRATSTSSSTPTSSGVAFAVN 99
QY 127 GGASGNGVTRYVDCCKASCSWPGKAN-----VSPYKSCTNKGCVTALSSNAOS 176
DB 100 IAGDEGCTTD--GTCVTSKYVPPLKNTFTGSNNYPDGIGQMOHFVNEGMITI----- 149
QY 177 GCNGGNSYMCNDNQMAVNDNLAYGFAPAAALISGGESESRWCCSCFILTPTSTVSAAKKNV 236
DB 150 -----FLRLPVGMWLYVNNNL-----GGLDSTISISKYQLVYQGLSLGAYCIV 192
QY 237 QVTN---TGGLSS--TGAHFDLQMPGGVGIFNGCSSQMGALNDGM 279
DB 193 DIHNVARWNGGIIIGCGGPTNAQF-----TSLMSQLASKVASQSRWV 233

```

ID	GNB	FUSOX	STANDARD	PR	462 AA
AC	P46235				
DT	01-NOV-1995	(Rel. 32, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	01-NOV-1995	(Rel. 32, Last annotation update)			
DE	Putative endoglucanase type B precursor (EC 3.2.1.4)	(Endo-1,4-beta-glucanase) (Cellulase)			
OS	Fusarium oxysporum				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OC	Hypocreales; mitosporic Hypocreales; Fusarium				
ON	NCBI_TaxID=5507;				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=95047531; PubMed=7959045;				
RA	Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,				
RA	Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;				
RT	"The use of conserved cellulase family-specific sequences to clone				
RT	cellulase homologue cDNAs from Fusarium oxysporum.";				
RL	Gene 150.163-167(1994).				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic				
CC	linkages in cellulose.				
CC	-1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).				
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL				
CC	HYDROLASSES).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on ways				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/annouce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/annouce/or_send_an_email_to_license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; L29377; AAA65585.1; -				
DR	HSSP; P07987; 1CB2.				
DR	InterPro; IPR000254; CBD_fungal.				
DR	InterPro; IPR001524; GH_6.				
DR	Pfam; PF00734; CBM_1; 1.				
DR	Pfam; PF01341; Glyco_hydro_6; 1.				
DR	PRINTS; PR00733; GLHYDRLASE6.				
DR	ProDom; PD001821; CBD_fungal; 1.				
DR	ProDom; PD003733; GH_6; 1.				
DR	SMART; SM00236; ECBD; 1.				
DR	PROSITE; PS00562; CBD_FUNGAL; 1.				
DR	PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.				
DR	PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.				
KW	Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.				
FT	SIGNAL	1	16	POTENTIAL.	
FT	CHAIN	17	462	POTENTIAL.	
FT	DOMAIN	17	65	POTENTIAL.	
FT	DOMAIN	66	99	CELLULOSE-BINDING.	
FT	DOMAIN	100	462	CELLULOSE-BINDING.	
FT	ACT_SITE	190	190	CATALYTIC.	
FT	ACT_SITE	236	236	PROTON DONOR (BY SIMILARITY).	
FT	ACT_SITE	416	416	NUCLEOPHILE (BY SIMILARITY).	
FT	DISULFID	33	50	BY SIMILARITY.	
FT	DISULFID	44	60	BY SIMILARITY.	
FT	DISULFID	191	250	BY SIMILARITY.	
FT	DISULFID	383	430	BY SIMILARITY.	
FT	CARBOHYD	37	37	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	223	223	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	272	272	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	317	317	N-LINKED (GLCNAC. . .) (POTENTIAL).	
SO	SEQUENCE	462 AA;	49207 MW;	E25B2F5B828B637F CRC64;	
QY	Query Match	9.7%;	Score 179;	DB 1;	Length 462;
QY	Best Local Similarity	27.4%;	Pred. No. 3.9e-05;		
DB	Matches	59;	Conservative 21;	Mismatches 61;	Indels 74;
DB					Gaps 9;

```

DR PROSITE; PS00562; CBD_FUNGAL; 1. 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
KM SIGNAL 1 20
FT CHAIN 21 438
FT DOMAIN 21 59
FT DOMAIN 60 87
FT DOMAIN 88 438
FT ACT_SITE 215 215
FT ACT_SITE 393 393
FT DISULFID 28 45
FT DISULFID 39 55
FT DISULFID 170 229
FT DISULFID 360 407
FT DISULFID 133 133
FT VARIANT 152 152
FT VARIANT 244 244
FT VARIANT 248 248
FT VARIANT 398 398
SQ SEQUENCE 438 AA; 46209 MW; 002C973544893794 CRC64;

Query Match 9.4%; Score 173; DB 1; Length 438;
Best Local Similarity 39.8%; Pred. No. 9.1e-05;
Matches 45; Conservative 10; Mismatches 36; Indels 22; Gaps 4;

OY 9 SALALALGTEMASAECSLKLYQCCGCKMNGTCESGTCVSNDDYSOCLEPGSSGN 68
:||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 AALALMSLVPGVQAO-SPGWGCGGNGMTGFTTCASSSTCYKONDPIFSQCLPPN---- 60

OY 69 KSESAAHKTTAAHKTTTAAHKTTTAAHKTTTAAKTTTAAKSTPNSSSSSSGK 121
:||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 -----APSTTT---QPGTTTPA---TTTSGGTGTPSGAGNPYTK 95

RESULT 9
GUNF_FUSOX
ID GUNF_FUSOX STANDARD; PRT; 385 AA.
AC P46239;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Putative endoglucanase type F precursor (EC 3.2.1.4) (Endo-1,4-beta-
glucanase) (Cellulase).
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
NCBI_Taxid=5507;
RN [1]
RP MEDLINE=95047511; PubMed=7959045;
RX Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,
RA Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;
RT "The use of conserved cellulase family-specific sequences to clone
celulase homologue cDNAs from Fusarium oxysporum.";
RL Gene 150:163-167(1994).
CC CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
links in cellulose.
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
CC -----
CC EMBL; L29380; AAA65588.1; -
CC DR HSSP; P56588; IBG4.
CC DR InterPro; IPR000254; CBD_fungal.

```



DR InterPro: IPR001000; Glyco\_hydro.10.  
 DR Pfam: PF00031; Glyco\_hydro\_10; 1.  
 DR Pfam: PF00734; CBM\_1; 1.  
 DR PRINTS: PR00134; GLHYDRLASE10.  
 DR ProDom: PD001821; CBD\_fungal; 1.  
 DR SMART: SM00236; fCBD; 1.  
 DR PROSITE: PS00562; CBD\_FUNGAL; 1.  
 DR PROSITE: PS00591; GLYCOSYL\_HYDROL\_F10; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 385  
 FT DOMAIN 25 52  
 FT DOMAIN 53 84  
 FT DOMAIN 85 385  
 FT ACT\_SITE 209 209  
 FT ACT\_SITE 321 321  
 FT ACT\_SITE 385 AA; 41225 MW; B3C3807C0D3C0EC CRC64;  
 SQ SEQUENCE 385 AA; 41225 MW; B3C3807C0D3C0EC CRC64;  
 Query Match 8.4%; Score 153.5; DB 1; Length 385;  
 Best Local Similarity 25.0%; Pred. No. 0.0016;  
 Matches 46; Conservative 20; Mismatches 59; Indels 59; Gaps 6;  
 QY 9 SALLALALGTEMASAECSKLYGCGKMNNGPTCCSGSTCKVSNDDYSGCLPSSGSGN 68  
 Db 5 SYLLALAPVSALAQA-----PIWGCGGNGMTGATTCASGLKCEKINDWYQCVP--GSGG- 58  
 QY 69 KSESASHKTTTAAHKTTTAAHKTTTAAKTITVAKASTPSSSSSSSGKYSVSGG 128  
 Db 59 -----SEPCPSSTGGGTPPTGPN 78  
 QY 129 ASGNGVTRTYWDCCKASCSPWKANVSPVSKCNKGVTLSDSNMGCGNGSNYMCND 188  
 Db 79 SGGTGIDAKF---KAK---GKQYFGTEIDHHLNPLINLYKQFG-----QYTCEN 125  
 QY 189 NQFW 192  
 Db 126 SMKW 129  
 RESULT 10  
 G168\_PARP STANDARD; PRT; 2704 AA.  
 ID G168\_PARP  
 AC P17053;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-AUG-1990 (Rel. 15, Last annotation update)  
 DE 168G surface protein precursor.  
 GN 168G.  
 OS Parametium primaurelia.  
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Penicillida;  
 OC Parametium.  
 OC NCBI\_TaxID=5886;  
 RX MEDLINE=90172419; PubMed=2308165;  
 RA Prati A.;  
 RT "Conserved sequences flank variable tandem repeats in two alleles of  
 the G surface protein of Parametium primaurelia.";  
 RL J. Mol. Biol. 211:521-535(1990).  
 CC -1- FUNCTION: THIS PROTEIN IS THE SURFACE ANTIGEN OR IMMOBILIZATION  
 ANTIGEN OF PARAMETIUM PRIMAURELIA.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- MISCELLANEOUS: IT HAS INTERNAL HOMOLOGIES AND A HIGHLY PERIODIC  
 STRUCTURE WITH 37 PERIODS OF ABOUT 75 RESIDUES, EACH PERIOD  
 CONTAINING 8 CYSTEINES, EXCEPT FOR FOUR HALF PERIODS. A VARIABLE  
 PART OF 475 RESIDUES COMPRISES 4 ALMOST IDENTICAL PERIODS IN THE  
 MIDDLE OF THE PROTEIN.  
 CC -1- MISCELLANEOUS: EXPRESSION OF G PROTEIN OCCURS AT LOW TEMPERATURES  
 (14-32 DEGREES CELSIUS).  
 CC -1- SIMILARITY: 98% TO THE ALLELIC FORM 156G PROTEIN (AC P13837) IN  
 CC PERIODIC STRUCTURE AND 80% IN THE VARIABLE DOMAIN IN THE MIDDLE  
 CC OF THE PROTEIN.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X52133; CA36378.1; -.  
 DR PIR: S09118; S09118.  
 DR InterPro: IPR002895; Parametium SA.  
 DR Pfam: PF01508; Parametium SA; 34.  
 DR Signal; Repeat; Antigen; Membrane; GPI-anchor.  
 FT SIGNAL 1 20  
 FT CHAIN 21 2704  
 FT DOMAIN 106 2560  
 FT DOMAIN 1060 1424  
 FT SEQUENCE 2704 AA; 27875 MW; 40EAD0A0B18EE2119 CRC64;  
 SQ SEQUENCE 2704 AA; 27875 MW; 40EAD0A0B18EE2119 CRC64;  
 Query Match 8.2%; Score 150; DB 1; Length 2704;  
 Best Local Similarity 22.3%; Pred. No. 0.017;  
 Matches 82; Conservative 33; Mismatches 148; Indels 104; Gaps 17;  
 QY 30 YGCGGKMNNGPTCC-----SGSTC-----KVSND-----YYSOCLPSSGSGN 68  
 Db 2271 YTSCKSLMNNDSCSKWISNCKTTGNSNCVGTLCSETNTDGCYTGDAICQSVPALN 2330  
 QY 69 KSESASHKTTTAAHKTTTAAHKTTTAAKTITVAKASTPSSSSSSSGKYSVSGG 124  
 Db 2331 SSDPKVKCKRYTSCAFAFTT--HSDQGLASSKCTTGTCIALSSCSYTAQAGCYFND 2388  
 QY 125 -----VSGASGNGVTRTYWDCCKASC-----SWPKANVSPVSKCNKGVTL-- 169  
 Db 2389 KGTLYTSYGVTSTGICT--WDTTSSSCRDQSCADLTGTHATCSQSLSTSDGTCLLK 2446  
 QY 170 -----SDSNMGCGNG--GNSYMCNDNQPMAVNDNLAYG--FAAAAI SGGESRMC----- 216  
 Db 2447 GACTSYTTOCTTAVGSDGACYWELASRYNNNTAKRLLTADIONGTAINVCVALST 2506  
 QY 217 -----CSCF--ELFTSTSVAGKRVVQVNTGDLGSSGTGAFDLPFGG 261  
 Db 2507 CYSNGTACIPKANCSTYTSKICNCGGLDGLCVFPQSTATGAAGTGTCA----- 2556  
 QY 262 VGIPIFGCSSQWCAPND-----GW--GSRYGISASDCS---SLPSALQAGCK 304  
 Db 2557 --LMTACTV---ANNQDQCAQARDRCGWTASGTGATVASKCATHTCATNQTATGACT 2611  
 QY 305 WRFNWF 311  
 Db 2612 RFLNWDK 2618  
 RESULT 11  
 SIM1\_YEAST STANDARD; PRT; 475 AA.  
 ID SIM1\_YEAST  
 AC P40472;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE SIM1 protein precursor.  
 GN SIM1 OR PB3 OR YIL123W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_TaxID=4932;  
 RX MEDLINE=90172419; PubMed=2308165;  
 RA Prati A.;  
 RT "Conserved sequences flank variable tandem repeats in two alleles of  
 the G surface protein of Parametium primaurelia.";  
 RL J. Mol. Biol. 211:521-535(1990).  
 CC -1- FUNCTION: THIS PROTEIN IS THE SURFACE ANTIGEN OR IMMOBILIZATION  
 ANTIGEN OF PARAMETIUM PRIMAURELIA.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- MISCELLANEOUS: IT HAS INTERNAL HOMOLOGIES AND A HIGHLY PERIODIC  
 STRUCTURE WITH 37 PERIODS OF ABOUT 75 RESIDUES, EACH PERIOD  
 CONTAINING 8 CYSTEINES, EXCEPT FOR FOUR HALF PERIODS. A VARIABLE  
 PART OF 475 RESIDUES COMPRISES 4 ALMOST IDENTICAL PERIODS IN THE  
 MIDDLE OF THE PROTEIN.  
 CC -1- MISCELLANEOUS: EXPRESSION OF G PROTEIN OCCURS AT LOW TEMPERATURES  
 (14-32 DEGREES CELSIUS).  
 CC -1- SIMILARITY: 98% TO THE ALLELIC FORM 156G PROTEIN (AC P13837) IN  
 CC PERIODIC STRUCTURE AND 80% IN THE VARIABLE DOMAIN IN THE MIDDLE  
 CC OF THE PROTEIN.

RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.,  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP IDENTIFICATION.  
 RX MEDLINE=86155611; PubMed=8574583;  
 RA Dahmann C., Diffley J.F.X., Nasmyth K.A.,  
 RT "s-phase-promoting cyclin-dependent kinases prevent re-replication by  
 RT inhibiting the transition of replication origins to a pre-replicative  
 RT state.";  
 RL Curr. Biol. 5:1257-1269 (1995).  
 CC -1- FUNCTION: REQUIRED FOR THE MAINTENANCE OF THE CLB5 KINASE  
 CC ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO THE SUN FAMILY.  
 CC -1- SIMILARITY: BELONGS TO THE SUN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: Z46833; CAAB6869.1; -  
 DR SGD: S0001385; SIM1.  
 KW Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 475 SIM1 PROTEIN.  
 FT DOMAIN 58 112 ALA-RICH.  
 FT DOMAIN 80 203 SER-RICH.  
 FT CARBOHYD 422 422 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 475 AA: 48070 MW: 8780 PDB: 8478612P CRC64;  
 Query Match 8.1%; Score 149.5; DB 1; Length 475;  
 Best Local Similarity 20.4%; Pred. No. 0.0035;  
 Matches 71; Conservative 47; Mismatches 131; Indels 99; Gaps 13;  
 Oy 5 TTASALLALALGTEMASAGCSKLYGCGGGMNMGPTCCBSGSGTCKVSNQDYSCLPSPG 64  
 Db 79 TSSAAGTAASINAVSALAKREKISDAAA---SATSTSGGASSSSSSSSASTLESS 134  
 Oy 65 SSGNKSSESAAHKT---TTAAHKTTTAAHKTTPAPKTTTAAKASTPSSSSSSSGK 121  
 Db 135 SVSSSSSEAAPTSTVSTTSATQSSASATKSTSTSTSTSTSTSTSTSTSTSTSTSSSSSS 194  
 Oy 122 YSANSQGA---SNGVYTRRY- 139  
 Db 195 SSSSSGSGSIYGLDLPFGSPSEKFPDGTIPCDKFPSPGGGVSTIDMIGEGMSGVENDTDS 254  
 Oy 140 ---DC---CKASGMPGKANVSPVKSCKNDGVTALSDSNAQSG--CNGANSYMCNDN 189  
 Db 255 TGSGCKESGYSYSCQ-PEMSKTQMPSDO-----PSDGRSVGGLCKRGYLYRSTTD 305  
 Oy 190 QPMVAVNDLAVGFAPAAAISSGGSERWCCSCEFLFTSTSVAG-KKVVQVNTNGDL--- 245  
 Db 306 ADYLCEWGEAAYVVSXLSKG-----VALCRIDYPTENVMVPTVYEGGSSSLPL 354  
 Oy 246 -----GSSTGAHFDLQMPGGVGFINGSSSQMGAPNDWGS 281  
 Db 355 TVVDODYFTWEGKTSAY-YYNNAAGVSVDGCI-WGTSGSGSIGN 398  
 RESULT 12  
 GUXI ASPAC STANDARD; PRT; 540 AA.  
 ID GUXI ASPAC  
 AC 059843;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellobiohydrolase I) (1,4-  
 DE beta-cellulohydrolase I) (beta-glucanocellulohydrolase I).  
 GN CBHI.

OS Aspergillus aculeatus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eucristales; Trichocommataceae; mitosporic Trichocommataceae; Aspergillus.  
 OX NCBI\_TaxID=5053;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F-50;  
 RA Takada G., Kawaguchi T., Sumitani J., Arai M.,  
 RT "Cloning, nucleotide sequence, and transcriptional analysis of  
 RT Aspergillus aculeatus No.F-50 cellobiohydrolase I (cbhi) gene.";  
 RL J. Ferment. Bioeng. 85:1-9 (1998)  
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE  
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:  
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS,  
 CC (2) EXOCELLULOSE HYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE  
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;  
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYSE THE CELLOBIOSE AND OTHER  
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.  
 CC CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
 CC in cellulose and cellobiose, releasing cellobiose from the non-  
 CC reducing ends of the chains.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL  
 CC HYDROLASES). CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE-BINDING DOMAIN (CBD).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AB002821; BAA25183.1; -  
 DR HSSP: P00725; 1A26.  
 DR InterPro: IPR000254; CBD\_fungal.  
 DR InterPro: IPR001722; GH\_7.  
 DR Pfam: PF00734; CBM\_1; 1.  
 DR Pfam: PF00840; Glyco\_hydro\_7; 1.  
 DR Prodom: PD001821; CBD\_fungal; 1.  
 DR Prodom: PD186135; GH\_7; 1.  
 DR SMART: SM00236; fcbd\_1.  
 DR PROSITE: PS00562; CBD\_FUNGAL; FALSE NEG.  
 DR Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.  
 KW SIGNAL 1 22  
 FT CHAIN 23 540  
 FT DOMAIN 23 459  
 FT DOMAIN 460 505  
 FT DOMAIN 506 540  
 FT ACT\_SITE 234 234  
 FT ACT\_SITE 239 239  
 FT DISULFID 513 529  
 FT DISULFID 524 539  
 FT CARBOHYD 434 434  
 FT CARBOHYD 458 458  
 SO SEQUENCE 540 AA: 57099 MW: 0027FD28DP194D0 CRC64;  
 Query Match 8.1%; Score 148.5; DB 1; Length 540;  
 Best Local Similarity 59.1%; Pred. No. 0.0046;  
 Matches 26; Conservative 5; Mismatches 12; Indels 1; Gaps 1;  
 Oy 18 TENASAAECSKLYGCGGKRWNGPTCCGSGSTCKVSNQDYSCQL 61  
 Db 498 TTSSSTNVAQLYGGCGGGMWPTTCASG-TLCKNDYISQCL 540  
 RESULT 13  
 PSBP\_PORPU STANDARD; PRT; 210 AA.  
 ID PSBP\_PORPU  
 AC PS0272;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

```

RP SEQUENCE FROM N.A.
RX STRAIN=AX4;
RA MEDLINE=96081966; PubMed=7499424;
RA Souza G.M., Hirai J., Mehra D.P., Freeze H.H.;
RT "Identification of two novel Dictyostelium discoideum cysteine
RT proteinases that carry N-acetylglycosamine-1-P-modification.";
RL J. Biol. Chem. 270:28938-28945(1995).
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- DEVELOPMENTAL STAGE: PRESENT IN THE VEGETATIVE PHASE AND DECREASES
CC WITH THE START DEVELOPMENT, REAPPEARS IN LOW LEVELS WHEN THE
CC FRUITING BODY IS FORMED.
CC -1- PPM: PHOSPHOGLYCOSYLATED, CONTAINS GLCNAC-ALPHA-1-P-SER RESIDUES
CC AND FUCOSE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J36204; AAA92019.1; -.
DR HSSP: P07711; ICUL.
DR MEROPS: C01.081; -.
DR Dictydb: DD01060; cpyD.
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR000169; Shprot_acsite.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS: PR00705; PAPAIN.
DR Prodom: PD000158; Peptidase_C1; 2.
DR PROSITE: PS00139; THIOI_PROTEASE_CYS_1.
DR PROSITE: PS00639; THIOI_PROTEASE_HIS_1.
DR PROSITE: PS00640; THIOI_PROTEASE ASN_1.
KW Hydroxylase; Thiol protease; Lysozyme; Zymogen; Glycoprotein;
KW phosphorylation; Signal; Repeat.
KM SIGNAL 1 17 POTENTIAL.
FT FT PROPEP 18 111 ACTIVATION PEPTIDE (POTENTIAL).
FT FT CHAIN 112 442 CYSTEINE PEPTIDINASE 4.
FT FT DOMAIN 286 296 POLY-SER.
FT FT DOMAIN 299 304 POLY-SER.
FT FT DOMAIN 310 315 POLY-SER.
FT FT DOMAIN 319 323 POLY-SER.
FT FT DOMAIN 327 332 POLY-SER.
FT FT ACT_SITE 135 135 BY SIMILARITY.
FT FT ACT_SITE 277 277 BY SIMILARITY.
FT FT ACT_SITE 406 406 BY SIMILARITY.
FT FT DISULFID 132 178 BY SIMILARITY.
FT FT DISULFID 169 212 BY SIMILARITY.
FT FT DISULFID 270 428 BY SIMILARITY.
FT FT CARBOHYD 228 228 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 442 AA; 45690 MW; F64FA816B9CF2434 CRC64;
Query Match 8.0%; Score 147.5; DB 1; Length 442;
Best Local Similarity 25.6%; Pred. No. 0.0044;
Matches 80; Conservative 28; Mismatches 115; Indels 89; Gaps 14
QY 14 LALGT--EMSAAE---CSKLYGQ--CGKK-----NNMG-----PTCCSG 47
DB 150 IASGPKDLVLSLSDNLIDCKSKSYGNNGCEGLMTLGFEYIIINNKGIDTESSPYTAABG 209
QY 48 STCKVNDYYVS-----QCPLSGSSGNKSSESAAHKTTTA--AHKTTTAAHKKTTTAPA 99
DB 210 KECKFKTSNIGAQIVSYONYNTSSGSBEALQSASNNAPAVVALDASNESPOLVESGIYEBA 269
QY 100 KKTT-----TVAKASTPSNSSSSSSSGRYAVSGAAGNVTRYMDCKRASCWPGKA 152
DB 270 CTPTQLDHGVLVWGSGSSSSSSSGSSSGKSSSSS---STGGKTSSSSSSSGKRASSSSSGVA 326
QY 153 NVSPSPVASCMNDGYTALLSDSNAGSCGNCNGNSYMCNDNPMAVNVDLIAIGFAAAALSQGE 212

```

Db 327 SSSS---SSGKTSAASTSGSGSGSGSQ-----SGQSTGSG 362

Qy 213 SRRCSCFELFTSTVAGKMKVQVNTGDDLGSSGTAHFDLWPGGIVGIFNGCSSQW 272

Db 363 S-----GGTASAGQ---ASAGSGSGSGSGSGSGS-----SGSGGAAYASSGNY 401

Qy 273 GAPNDGWSRYG 284

Db 402 WIVNWSMGTSWG 413

## RESULT 15

GUNS TRIRE STANDARD; PRT; 242 AA.

AC P4317;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)  
DE (Cellulase V) (EG V).  
GN EGL5.  
OS Trichoderma reesei (Hypocrea jecorina).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Hypocreales; Hypocreaceae; Hypocrea.  
OX NCBI\_TaxID=51453;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CM9414 / Rut C-30;  
RX MEDLINE=95075308; PubMed=7984103;  
RA Saloheimo A., Henriksa B., Hoffren A.-M., Telemann O., Penttilae M.;  
RT "A novel, small endoglucanase gene, egl5, from Trichoderma reesei  
isolated by expression in yeast."  
RL Mol. Microbiol. 13:219-228(1994).  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose.  
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL  
HYDROLASES).  
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; Z33381; CA83846.1; -.  
DR HSSP; P00725; ZCBH.  
DR InterPro; IPR000254; CBD\_fungal.  
DR InterPro; IPR000334; GH\_45.  
DR Pfam; PF00734; CBM\_1; 1.  
DR Pfam; PF02015; Glyco\_hydro\_45; 1.  
DR ProDom; PD001821; CBD\_fungal; 1.  
DR SMART; SM00236; FCB; 1.  
DR PROSITE; PS00562; CBD\_FUNGAL; 1.  
DR PROSITE; PS50842; EXPANSIN\_EG45; 1.  
DR PROSITE; PS01140; GLYCOSYL\_HYDROL\_F45; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1  
FT CHAIN 18 242  
FT ACT\_SITE 27 27 ENDOGLUCANASE V.  
FT ACT\_SITE 134 134 NUCLEOPHILE (BY SIMILARITY).  
FT DOMAIN 18 182 PROTON DONOR (BY SIMILARITY).  
FT DOMAIN 183 205 CATALYTIC.  
FT DOMAIN 206 242 PRO/SER-RICH (LINKER).  
FT CARBOHYD 182 182 CELLULOSE-BINDING (BY SIMILARITY).  
FT DISULFID 213 230 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DISULFID 224 240 BY SIMILARITY.  
SQ SEQUENCE 242 AA; 24411 MW; CC033FC51326C71D CRC64;

Query Match 8.0%; Score 147; DB 1; Length 242;  
Best Local Similarity 67.6%; Pred. No. 0.0027;

Matches 23; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Qy 29 LVGQCGGKXNMNPTCCSGSTCKVNSNDYYSQCLP 62  
Db 209 LVGQCGGAGWTGPTTCQAPETCKVQNMYSQCLP 242

Search completed: June 18, 2003, 15:32:01  
Job time : 10.8656 secs